

2e Rencontre Fribourgeoise de la Linguistique sur Corpus, Fribourg-en-Brigau, septembre 2003:—Du 11 au 14 septembre 2003 eut lieu, à l'Université «Albert Ludwig» de Fribourg-en-Brigau (Allemagne), la 2e Rencontre Fribourgeoise de la Linguistique sur Corpus Appliquée aux Langues Romanes (2nd Freiburg Workshop on Romance Corpus Linguistics). Ce colloque faisait suite à une première rencontre organisée en 2000 et dont le sujet principal était l'analyse du langage parlé à travers les corpus et les bases de données.³ Le deuxième colloque prenait une autre perspective et se penchait sur la thématique de l'utilité des corpus électroniques et des banques de données pour la linguistique historique.

Il est évident que la description et l'explication du changement linguistique diachronique demande l'examen comparatif et contrastif d'un grand nombre de témoignages textuels très divers que le chercheur doit organiser et structurer afin d'arriver à une analyse exacte et saisissable. Dans ce sens, toute linguistique historique —indépendamment de l'approche privilégiée, intra-langue ou comparative— est une linguistique sur corpus, et Christiane Marchello-Nizia (Paris), dans sa conférence plénière intitulée «Pour une linguistique historique outillée», a parfaitement illustré ce fait par rapport à l'étude de l'ancien et du moyen français. Or, le but de la rencontre fribourgeoise était de savoir dans quelle mesure les méthodes et les techniques de la linguistique sur corpus au sens moderne du mot, c'est à dire d'une linguistique de texte assistée par ordinateur, peuvent être mises à profit dans la recherche sur le changement diachronique des langues et quels résultats innovateurs peuvent résulter de l'application de ces outils informatiques dans le domaine de la linguistique historique.

La colloque a rassemblé dans la capitale de la Forêt Noire un total de 85 chercheurs en provenance de 17 pays d'Europe et d'Outre-mer, qui se sont organisés en deux sections, dont la première était consacrée aux aspects techniques de la constitution et de l'utilisation des corpus historiques électroniques et aux questions méthodiques de base. Parmi les projets de corpus présentés dans cette section figuraient la «Base du Français Médiéval», élaborée par Christiane Marchello-Nizia (Paris), Sophie Prévost (Paris), Serge Heiden (Lyon) *et al.*; l'impressionnant «Corpus del español» d'une extension de 100 millions de mots entièrement étiquetés, élaboré par Mark Davies (Provo) et librement accessible par internet; le «Corpus del Gallego Actual», en élaboration sous la direction de Marisol López (Santiago de Compostela) et dont la perspective diachronique vise, de par les types de données intégrées, les changements dits «à court terme», qui sont en train de se dérouler ou qui se sont manifestés parmi les générations de locuteurs les plus récentes; ou enfin le «Corpus de Diatopia Diacrònica de la Llengua Catalana», présenté par M. Àngels Massip et Joan Veny (Barcelona). La seconde section du colloque était ouverte à la présentation de recherches sur la diachronie des langues romanes qui avaient, comme base empirique, un corpus constitué à ce propos ou qui utilisaient les corpus et les bases de données publiquement disponibles. Cette section offrait une grande variété d'approches, de sujets et de périodes historiques traités, et cela par rapport à pratiquement toutes les langues romanes (au sens vaste), à commencer par le latin archaïque et jusqu'aux langues créoles à base lexicale romane et aux variétés diasystémiques les plus récentes et les plus innovatrices comme, par exemple, les nouveaux «français urbains» d'Afrique.

Les conclusions tirées après cette 2e Rencontre Fribourgeoise de la Linguistique sur Corpus étaient multiples et fort instructives; elles seront développées dans le volume d'actes du colloque dont la publication est prévue pour 2005. Parmi les aspects les plus frappants, on peut certainement mentionner la constatation que les problèmes posés par la numérisation et l'ex-

3. Les actes de cette première rencontre sont désormais disponibles: Claus D. Pusch / Wolfgang Raible (éds.), *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache / Romance Corpus Linguistics: Corpora and Spoken Language*, Tübingen: Narr, 2002 (ScriptOra!ia; 126).

exploitation électronique du langage parlé, étudiés lors du colloque de 2000, ressemblent à maints égards aux problèmes que pose la numérisation des textes anciens et leurs préparation pour l'analyse assistée par ordinateur: les textes anciens sont aussi riches —voire plus riches encore— en phénomènes de polymorphismes, de variantes (morphologiques ou graphiques) et d'ambiguïtés, ce qui pose évidemment problème à l'étiquetage et à l'analyse syntaxique automatique (ou, au moins, partiellement automatisée). Mais tandis qu'en cas de textes oraux, on peut, à la rigueur, avoir recours à une transcription orthographique proche du standard, une telle solution s'interdit dans le cas de documents anciens, où elle risquerait de rendre méconnaissable le texte de départ et d'aliéner son authenticité comme «document historique». Ainsi, la numérisation des textes anciens, tout en étant une approche technique très récente, constitue un nouveau défi pour une des disciplines les plus anciennes et les plus importantes de la romanistique, à savoir l'édition philologique et son inventaire méthodique.

Vu le succès rencontré par la 2e Rencontre Fribourgeoise de la Linguistique sur Corpus, les organisateurs se sont décidés à poursuivre cette série de colloques et à prévoir dès maintenant une troisième rencontre, consacrée cette fois-là au sujet «corpus et pragmatique», pour 2006. Pour des renseignements mis à jour continuellement, un site internet permanent est disponible à l'adresse <<http://www.corpora-romanica.net>>. [C. D. P.]